International Journal of Information Management 45 (2019) 83–94

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

A General framework for studying the evolution of the digital innovation ecosystem: The case of big data

Bongsug (Kevin) Chae

Department of Management, Kansas State University, United States

ARTICLE INFO

Keywords:
Digital innovation ecosystem
Computational research
Network science
Big data

ABSTRACT

This paper presents a general framework for studying the digital innovation ecosystem. The notion of complex networks offers a conceptual lens to analyze the emergence and evolution of a digital innovation ecosystem. The framework uses digital data and evolutionary community detection analysis for the empirical inquiry of the digital innovation landscape. The proposed framework is applied to the big data ecosystem. Data from Twitter, for a three year period, is processed and analyzed. This study reveals a large number of elements that are diverse in form and capacity, including organizations, concepts (e.g., #analytics, #iot), technologies (e.g., #hadoop), applications (e.g., #healthcare), infrastructures (e.g., #cloud), regulations, professional meetings and associations, tools, and knowledge. These elements and their communities have evolved in the big data ecosystem. The findings highlight the evolution of digital innovation by two mechanisms, variation and selective retention, which are nonlinear and often unpredictable. Implications are presented and potential ways to improve the proposed framework are discussed. The study aims to make both conceptual and methodological contributions to digital innovation research.

1. Introduction

Digital innovation research brings digital technologies into the foreground of innovation management (Fichman, Dos Santos, & Zheng, 2014; Nambisan, Lyytinen, Majchrzak, & Song, 2017; Yoo, Henfridsson, & Lyytinen, 2010; Yoo, Boland, Lyytinen, & Majchrzak, 2012). Past digital innovation research has focused on the implementation and adoption of information technologies, such as enterprise systems and knowledge management systems, by individuals and organizations (Fichman et al., 2014). Some recent studies draw our attention instead to digital artifacts or products themselves, including emerging digital technologies and digital infrastructures (Yoo et al., 2010). Many theoretical frameworks and research methods have been brought into digital innovation (Fichman et al., 2014). Recent commentaries (Nambisan et al., 2017; Yoo, 2015) have suggested exploring new theoretical lenses and research methods for digital innovation research, including the use of digital data, computational algorithms, and evolutionary ontology.

In this line, this paper presents a general framework for studying how a digital innovation evolves over time. For innovation research, the unit of analysis can be within organizations, across firm networks, and at macro level or within communities (Garud, Tuertscher, & Van de Ven, 2013). This study adopts the macro level analysis for digital innovation, while using the term "digital innovation ecosystem", to refer to a complex arrangement of technologies, methodologies, concepts, business application areas, organizations, and institutional contexts. From the study of organizations, this level would be considered organizational collectives (Chiles, Meyer, & Hench, 2004) or organizational fields, which are "the least familiar, yet the level of most significance" (Scott, 2008, p. 86).

The proposed framework borrows the notion of a complex network (or system) (Axelrod & Cohen, 2000; Holland, 2014) as a theoretical lens to study digital innovation as a network of heterogeneous elements, which changes over time through such evolutionary mechanisms as variation and selective retention. The framework also adopts the digital and computational research tradition (Yoo, 2012, 2015) as the research methodology, while using digital data as its data source, and borrowing network science techniques.

The proposed framework is applied to the macro-level inquiry of big data as the latest case of digital innovation. Large quantities of digital data related to big data have been collected using hashtag #bigdata from Twitter, one of the fastest-growing social media platforms. Twitter data from a two-month period in each year (2013, 2014, and 2015) is analyzed using the proposed framework. Two key questions this paper attempts to answer using the proposed framework are: What has constituted the ecosystem of big data? And How has that ecosystem

E-mail address: bchae@ksu.edu.

https://doi.org/10.1016/j.ijinfomgt.2018.10.023
Received 8 January 2018; Received in revised form 30 October 2018; Accepted 30 October 2018
0268-4012/

evolved over the years? Through the proposed framework and its application to big data as a digital innovation, this study attempts to make both conceptual and methodological contributions to innovation management research, in general, and digital innovation research, in particular.

2. Background

Ecosystem has been one of the focal concepts used in some recent studies where digital product/service innovation is recognized as a process of reconfiguring or recombining existing resources available in a product/service ecosystem (Hu, Huang, Zeng, & Zhang, 2016; Tiwana, 2015; Um, Yoo, & Wattal, 2015). Similarly, this study uses the term “digital innovation ecosystem”, referring to a network of heterogeneous elements, which co-evolve over time. Thus, it becomes important to identify such heterogeneous elements and to explain the emergence and evolution of the ecosystem or network of such elements.

Both industry and academia have shown considerable interest in and attention to the benefits big data and analytics can offer, and the investment in big data skills and resources has increased over past years (Gandomi & Haider, 2015; Goes, 2014; Gupta, Kar, Baabdullah, & Al-Khowaiter, 2018; Tambe, 2014; Yaqoob et al., 2016). Big data is particularly interesting as a digital innovation case because it is disruptive and transformative in the entire value chain of information (data, information, knowledge, decisions, actions) and even scholarly research (Abbasi, Sarker, & Chiang, 2016; de Camargo Fiorini, Roman Pais Seles, Chiappetta Jabbour, Barberio Mariano, & de Sousa Jabbour, 2018; Raguseo, 2018).

Big data as an innovation involves not just digital technologies and tools, but also knowledge, skills, concepts, organizations, and other social and institutional contexts. This disruptive innovation is not created in a vacuum. Big data innovation has evolved from the past (e.g., business intelligence, data mining, data warehousing) and combines new resources, such as analytical platforms (e.g., R, Python), computing architecture (e.g., high performance computing), data processing frameworks (e.g., Hadoop), infrastructure (e.g., cloud computing, large data centers), analytical talents, beliefs, methodologies, professional meetings, and institutions (e.g., regulations, privacy) (2015a, Chae, 2014; Hashem et al., 2015; Sagiroglu & Sinanc, 2013).

3. A proposed framework for digital innovation research

The proposed framework combines complex networks as a conceptual lens and computational research methodology for the empirical inquiry. A digital innovation ecosystem is viewed as a complex network of heterogeneous social and technical elements and, thus, changes in the ecosystem are characterized as an evolutionary process involving “variation and selective retention” (Campbell, 1960).

3.1. Complex network as a conceptual lens

According to the literature (Axelrod & Cohen, 2000; Holland, 1995; Kauffman, 1995), complex networks, both natural and artificial, are exhibiting some fundamental characteristics, such as hierarchical and modular structure and continual novelty creation (or emergence), through evolution. Complex networks are composed of heterogeneous elements, also being “diverse in both form and capability” (Holland, 1995). The networks are not designed, but rather emerge from the interaction of individual elements or agents, which are not necessarily limited to either social or technical elements. The elements are ‘located’ in the sense that some elements may be close in proximity and this location results in more interaction among those elements (Axelrod & Cohen, 2000). The elements with high interaction form a module (or community) and, in turn, modules form a higher module or system (Holland, 1995). Therefore, hierarchical modularity is universal for complex networks (Axelrod & Cohen, 2000; Holland, 1995).

Complex networks re-emerge over time, as there are changes in elements and local interactions. The elements are continually reorganizing and forming communities. Also, new elements can be imported from neighboring systems, as complex networks co-evolve with other networks (Kauffman, 1995). This results in the emergence or renewal of communities and a digital innovation ecosystem. This continual renewal is the hallmark of complex networks (Holland, 1995). This can be characterized as a process involving two evolutionary mechanisms, variation and selective retention (Axelrod & Cohen, 2000). Variation increases diversity in complex networks, while selective retention changes the centrality (or importance) of elements and communities in complex networks.

3.2. Network science techniques for the empirical inquiry

There has been a bourgeoning of studies exploring diverse types of complex networks – network science (Strogatz, 2001), where computational algorithms and data visualization techniques are being developed to find hidden patterns within complex networks. Two such categories of network science techniques—community detection and network evolution—are highly relevant in studying the digital innovation ecosystem as a complex network.

Modularity is commonly observed in complex networks (Girvan & Newman, 2002). A module or community is “a cohesive group of nodes that are connected more densely to each other than to the nodes in other communities” (Porter, Onnela, & Mucha, 2009, p. 1086). Network science offers techniques for detecting communities within complex networks (Fortunato, 2010). The evolution of complex networks can be understood by focusing on the changes in communities. This is dynamic community (or topic) detection. Such changes include growth, merging, birth, contraction, splitting, and death (Palla, Barabási, & Vicsek, 2007). In complex networks, variation occurs through birth, merging, and splitting, in communities, while growth, contraction, and death are the mechanisms for selective retention.

This dynamic community detection has been used to understand the evolutionary changes of complex systems, including collaboration networks, economic networks, online social networks, human cooperation networks, financial lending networks, and academic fields (Bech, Bergstrom, Rosvall, & Garratt, 2015; Fortunato & Hric, 2016; Holme, 2015; Rossetti & Cazabet, 2018). Methods and metrics include instant optimal detection (e.g., Jaccard index), temporal trade-off discovery (e.g., simulated annealing), and cross-time community detection (Rossetti & Cazabet, 2018).

3.3. A general framework

Drawn from the discussion in Sections 3.1 and 3.2, a general framework for studying the emergence and evolution of the digital innovation ecosystem is presented (Fig. 1). The proposed framework succeeds the recent approach to digital innovation as a recombinant process of resources (e.g., Yoo et al., 2012), while conceptualizing the digital innovation ecosystem as a complex network of heterogeneous elements. The framework includes the following stages:

- Digital innovation ecosystem as a complex network.
- Digital data as the source of research data.
- Coding digital data as network nodes and edges for semantic network.
- Applying community detection analysis.
- Analyzing temporal networks at time (t) and another time (t + 1) through an evolutionary perspective on digital innovation ecosystems.

(1) Complex network as a conceptual lens: Diverse elements, such as technologies, popular concepts (Wang, 2009), methodologies, and institutional contexts, come into play in the emergence and

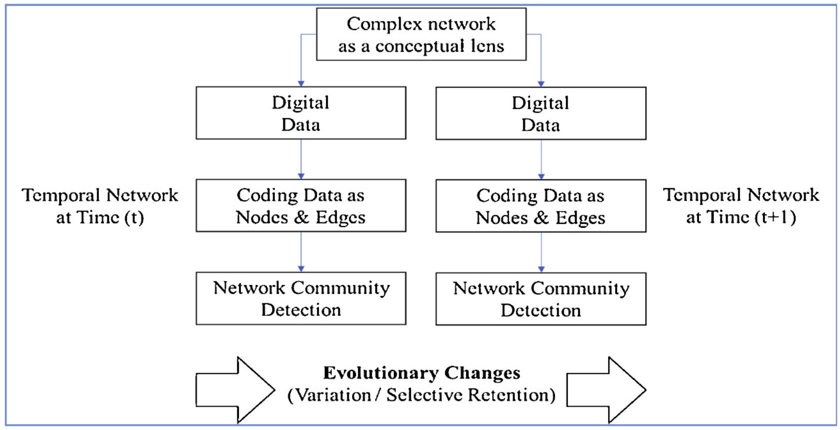

Fig. 1. A General Framework for Studying the Digital Innovation Ecosystem.

evolution of a digital innovation (e.g., big data). The proposed framework assumes that such elements interact dynamically and form communities in a digital innovation ecosystem. The notation of complex network is a helpful way of conceptualizing this phenomenon.

(2) Digital data as data source: The key to studying the digital innovation ecosystem as a complex network is identifying heterogeneous elements dynamically interacting and co-evolving. Since the goal is to study a digital innovation (e.g., big data, 3D printing) beyond individual organizations and an industry, the framework takes advantage of the availability of digital data at greater scale and algorithms and an evolutionary ontology (Yoo, 2015) in studying complex networks and in finding how novelties emerge (Tria, Loreto, Servedio, & Strogatz, 2014). In recent years, the use of digital data has become increasingly popular in information management studies (Aswani, Kar, Ilavarasan, & Dwivedi, 2018; Grover, Kar, & Davies, 2018; Grover, Kar, Dwivedi, & Janssen, 2018).

(3) Coding digital data as networks: The literature introduces a variety of techniques for text processing and analysis (Chae, 2015b; Chang, Ku, & Chen, 2017; Kapoor et al., 2018; Rathore, Kar, & Ilavarasan, 2017; Stieglitz, Mirbabaie, Ross, & Neuberger, 2018). Digital text data can be processed as either arrays of features or networks of co-appearing words. The network approach codes each word or token as a network node, and co-appearing words in the same documents are connected through network edges. This forms a semantic network (Steyvers & Tenenbaum, 2005). This network representation of digital data is especially suitable for studying the emergence and evolution of digital innovation ecosystems from the complexity lens introduced earlier.

(4) Community detection and evolutionary changes: The complexity network lens in Section 3.1 highlights the interaction of diverse elements with each other and how they self-organize into communities through variation and selective retention. The outcome is the changes in the digital innovation ecosystem. Network science offers many approaches and algorithms for mapping this evolutionary process. For example, the map equation (Bohlin, Edler, Lancichinetti, & Rosvall, 2014; Rosvall, Axelsson, & Bergstrom, 2009) provides a tool for community detection and network visualization.

Once communities are identified for different time frames (e.g., 2014, 2015) initially, the evolution can be analyzed through community similarity. Communities are similar where they share common nodes (Pan, Li, Liu, & Liang, 2010). Similarity metrics (e.g., Jaccard index) can measure the similarities and differences between communities at time t (e.g., 2014) and those at time t + 1 (2015). The idea of community similarity is linked into the complexity network lens, in that the boundary of a digital innovation ecosystem is in flux, existing and new elements are organizing into communities and, thus, communities share similarities and differences. The Jaccard index offers a way to quantify that.

Network visualization (e.g., alluvial diagrams) (Rosvall & Bergstrom, 2010) can then effectively map the details of the evolutionary changes. Variation can be observed when communities of nodes that were not seen in the network at time 1 (t), appear in the network in t + 1 (birth); the nodes in a community split into two or more communities (splitting); and the nodes from two or more communities form a new community (merging) (Pan et al., 2010). Selective retention involves death, growth, and contraction. A community can get larger (growth) or smaller (contraction) in terms of its number of nodes and their importance in the overall ecosystem, and also disappear (death). For example, this process makes certain elements (e.g., MapReduce, Apache Spark) and communities (e.g., Hadoop community) more or less popular than before.

4. Data and research method

4.1. Digital twitter as data source

Twitter is used as the source of digital data to trace diverse elements related to big data innovation. The social media platform provides unfiltered user-generated content for research (Grover, Kar, Davies et al., 2018; Grover, Kar, Dwivedi et al., 2018). Twitter as a social broadcasting technology (Shi, Rui, & Whinston, 2014) is popular for technology-related conversations, and IT professionals and businesses use Twitter extensively to stay current with industry changes and technology developments (Wang, Kuzmickaja, Stol, Abrahamsson, & Fitzgerald, 2014). These conversations are formed around Twitter hashtags, which evolve in Twitter.

4.2. Coding digital Twitter data as networks

Data collection focuses on capturing Twitter conversations related to big data. Twitter Streaming API was accessed during three periods (February-March 2013, May-June 2014, and September-October 2015). Every tweet containing the hashtag #bigdata was collected during these periods. Table 1 presents a summary of the data collected during the three periods.

Popular text processing techniques (e.g., lowercase, removing stopwords) were used to clean and transform data. This results in 11,101 unique hashtags in the 2013 network dataset, 17,450 in the 2014 network dataset, and 27,209 in the 2015 network dataset. The hashtag #bigdata was then removed from those network datasets, since that hashtag appears in every tweet. Some general hashtags (e.g.,

Table 1
Descriptive Statistics.

	Feb-Mar 2013	May-Jun 2014	Sep-Oct 2015
Tweets	174,703	321,059	1,175,895
English tweets (after removing non-English tweets)	136,816	246,574	996,968
Original tweets (after removing retweets)	80,765	129,128	346,696

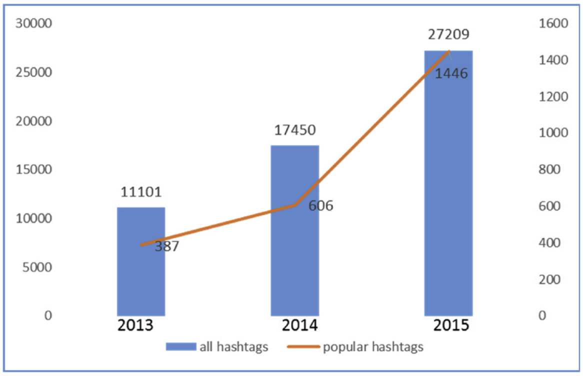

Fig. 2. Increases in heterogeneity of the big data ecosystem.

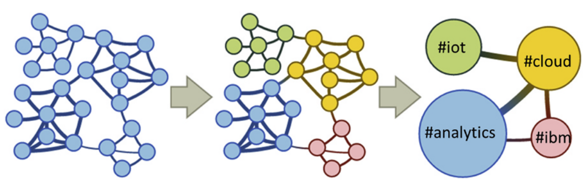

Fig. 3. An illustration of the community detection process (Adapted from www.mapequation.org/apps.html).

#news, #job) were also removed from the network datasets. Also, hashtags (e.g., #lawfirms) in fewer than 30 tweets in each data collection phase were removed.

Coding those popular hashtags and their connections as network data results in 387 nodes (or hashtags) and 38,078 edges in 2013, 606 nodes and 80,911 edges in 2014, and 1446 nodes and 610,262 edges in 2015 (Fig. 2). These hashtags are heterogeneous social and technical elements of the big data ecosystem. Over time, there is a significant increase in heterogeneity of elements in the ecosystem. Many new hashtags appear in 2014 (e.g., #deeplearning, #abdsc, #bdw2014) and even more in 2015 (e.g., #3dprinting, # refugeecrisis, #bgcon2015).

5. Analysis & results

5.1. Community detection analysis

In the map (Bohlin et al., 2014; Rosvall et al., 2009) (Fig. 3), communities are represented by different sizes of nodes, and the community label is derived from the most important hashtag (the hashtag with highest PageRank[1]). The assumption is that those nodes with high association with other nodes are important.

Table 2 shows the number of communities for each period of data analysis, community names labeled by most import hashtags, the number of hashtags (or nodes) per community, and the aggregated PageRank of all nodes in each community. On the other hand, Fig. 4 shows two types of maps: A whole map shows the communities and the links among them. A sub-network map depicts the details of each community. For illustration, one of the largest sub-networks (#iot) is included in the figure.

The number of communities differs significantly and increases over time. The results show 24, 42, and 70 communities in 2013, 2014, and 2015, respectively, indicating significant growth in terms of the boundary and complexity of the big data ecosystem. The communities differ in terms of their elements and sizes. For example, the "#iot" community is one of the smaller communities in 2013, with only seven hashtags (#iot, #m2m, #cisco, #energy, #internetofthings, #smartgrid, #industrialinternet), and low PageRank. This was a small, but highly specialized community (or a sub-field in big data ecosystem), focusing on Internet of Things. However, this small community has grown rapidly over time and become the largest community, with over 160 hashtags (e.g., #api, #devops, #java, #linux, #wearables, #openstack, #sdn, #fogcomputing) in 2015. Like the #iot community, the rest of the communities (e.g., #hadoop, #hr, #security) in the ecosystem represent sub-fields or key topical areas in the big data ecosystem, except #analytics community.

The "#analytics" community is the largest community in 2013 and 2014. Many popular concepts (e.g., #datascience), tools (e.g., #r, #python), technologies, organizations (e.g., #sas), knowledge (e.g., #statistics), methodologies (e.g., #agile), roles (e.g., #datascientist), application areas (e.g., #insurance), and issues (e.g., #privacy) have connections with #analytics. This makes #analytics the largest community, in which almost 40% of information flows (or PageRank) are concentrated. Over time, some of the elements split from this community and create their own communities (e.g., #aws, #privacy) or merge with other existing communities (e.g., #cloud) in 2014 and 2015.

5.2. Evolutionary changes: variation & selective retention

Fig. 5 represents the results of the Jaccard index-based community similarity in heatmaps. The heatmaps show that big data ecosystems have evolved between 2013 and 2015, and that communities between 2013 and 2014 (and between 2014 and 2015) share common nodes. For example, #iot community in 2013 is similar to three communities, #cloud, #sdn, and #energy, in 2014. Also, #cloud community in 2014 is very similar to #iot in 2015, in terms of their member nodes.

The map equation's alluvial diagrams were then used to map variation and selective retention, two mechanisms in the evolutionary process of the big data ecosystem. An alluvial diagram is created for each year, with communities (or blocks) ordered by size. Fig. 6 shows structural changes in the big data ecosystem over a period of time (2013, 2014, and 2015). Also, the figure highlights the evolution of one specific community (#iot) as an example. The alluvial diagrams reveal dynamic patterns of variation and selective retention in the big data ecosystem. The changes of elements and their connections lead to community-level changes, resulting in further changes in the ecosystem. Over time, existing elements have rearranged themselves, through merging and splitting, and new elements have appeared in the ecosystem. This has resulted in the emergence of new communities and the growth, contraction, and disappearance of existing communities.

Fig. 7 highlights variation, including birth, merging, and splitting. Communities, such as #climate, #agriculture, and #threat, emerged and led to diversity in the big data ecosystem. New communities initially emerged as small by size (less important). These communities (e.g., #agriculture) contain elements mostly imported from other

[1] PageRank is an algorithm originally developed by Google to determine the importance of websites (Page, Brin, Motwani, & Winograd, 1999).

Table 2
The results of the community detection analysis.

Community	2013			2014			2015		
	Central[a]	Counts[b]	PageRank[c]	Central	Counts	PageRank	Central	Counts	PageRank
1	#analytics	159	0.406	#analytics	233	0.311	#iot	166	0.413
2	#cloud	32	0.198	#cloud	38	0.230	#analytics	745	0.312
3	#hadoop	48	0.154	#privacy	30	0.086	#innovation	8	0.075
4	#healthcare	23	0.039	#hadoop	49	0.086	#hadoop	99	0.051
5	#security	18	0.035	#innovation	10	0.043	#cio	9	0.028
6	#linkeddata	15	0.035	#healthcare	38	0.035	#hr	23	0.021
7	#cxo	14	0.023	#security	19	0.024	#aws	27	0.018
8	#innovation	5	0.017	#mobile	8	0.022	#social	13	0.006
9	#enterprise	12	0.013	#business	14	0.019	#predictiveanalytics	9	0.005
10	#sap	9	0.011	#social	9	0.018	#mobility	12	0.005
11	#emc	4	0.010	#analysis	7	0.016	#growthhacking	22	0.005
12	#socbiz	4	0.009	#sdn	10	0.014	#research	8	0.004
13	#mdm	6	0.009	#hr	12	0.010	#finance	5	0.003
14	#hr	5	0.008	#highered	7	0.009	#genomics	15	0.003
15	#dataviz	3	0.007	#cybersecurity	5	0.009	#agtech	17	0.003
16	#aws	5	0.007	#sap	9	0.009	#emc	12	0.002
17	#iot	7	0.007	#opendata	10	0.007	#javascript	11	0.002
18	#sales	3	0.004	#dataanalytics	4	0.006	#customer	10	0.002
19	#aiim	5	0.003	#mdm	10	0.005	#hp	17	0.002
20	#cmos	3	0.002	#ecm	6	0.005	#google	8	0.002
21	#infosys	2	0.002	#club	5	0.005	#stock	6	0.002
22	#inspire13	2	0.001	#ceo	7	0.005	#datacenter	6	0.002
23	#netflix	2	0.001	#worldcup	6	0.004	#nfl	17	0.002
24	#actonsw	1	0	#database	5	0.004	#windows	5	0.002
25				#tiecon	3	0.002	#dataquality	12	0.002
26				#edtech	7	0.002	#developers	4	0.002
27				#web	4	0.002	#sap	7	0.002
28				#climate	5	0.002	#knowledge	3	0.002
29				#dataeast14	3	0.002	#threat	7	0.002
30				#datamanagement	5	0.002	#virtualreality	8	0.002
31				#toronto	3	0.001	#tungsten	4	0.001
32				#hpc	2	0.001	#infographics	4	0.001
33				#energy	3	0.001	#android	5	0.001
34				#learning	2	0.001	#oow15	6	0.001
35				#littledata	2	0.001	#internet	4	0.001
36				#integration	3	0.001	#gartnersym	10	0.001
37				#soccer	3	0.001	#infogov	4	0.001
38				#lean	2	0.001	#datarevolution	10	0.001
39				#bigquery	2	0.000	#hpc	3	0.001
40				#agriculture	2	0.000	#cities	4	0.001
41				#ldntechweek	2	0.000	#digitalstrategy	3	0.001
42				#bigdata2014	2	0.000	#competitive	3	0.001
43							#servers	6	0.001
44							#ford	5	0.001
45							#hcanalytics	3	0.000
46							#oil	3	0.000
47							#browserweb	4	0.000
48							#virtualization	4	0.000
49							#services	2	0.000
50							#qlik	2	0.000
51							#anonymization	3	0.000
52							#informationmanagement	3	0.000
53							#csharp	4	0.000
54							#kpi	2	0.000
55							#howto	2	0.000
56							#photography	3	0.000
57							#disruptive	3	0.000
58							#industry4	2	0.000
59							#facts	2	0.000
60							#casestudy	2	0.000
61							#disruptivetech	2	0.000
62							#telcobigdataasia	2	0.000
63							#govdata	2	0.000
64							#watsonanalytics	2	0.000
65							#ipaas	2	0.000
66							#migrants	2	0.000
67							#biztravel	2	0.000
68							#cannabis	2	0.000
69							#uncommons	2	0.000
70							#keynote	2	0.000

[a] Hashtag with highest PageRank.
[b] Number of hashtags in each community.

[c] Aggregated PageRank of all nodes in each community.

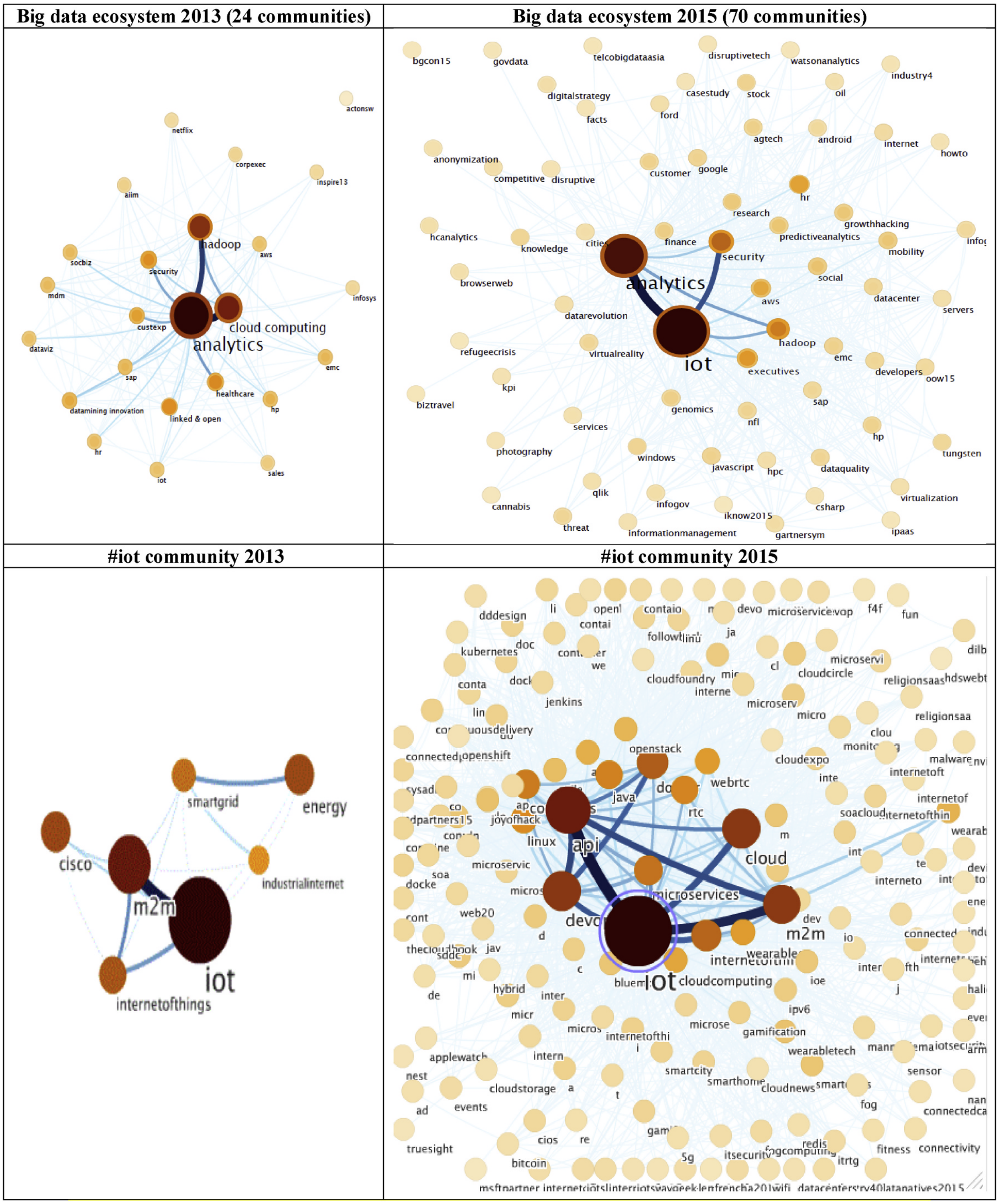

Fig. 4. The "map" view of changes in the big data ecosystem and one exemplar community (#iot).

ecosystems (e.g., ecosystem of agriculture), indicating the co-evolution of ecosystems. This may indicate the emergence of big data applications in the field of agriculture. #threat is another example of a new community in 2014, including #dhs (Dept. of Homeland Security), #isao (Information Sharing and Analysis Organizations), #ciso (Chief Information Security Officer), #isac (Information Sharing and Analysis Center), and other elements. The alluvial diagrams of 2013 and 2015 show the emergence of many new communities through birth. Communities (as well as the ecosystem itself) also emerge at t + 1 from communities merging and/or splitting at t. Organizations (e.g., #talend), concepts (e.g., #datalake), tools (e.g., #yarn), techniques (e.g., #spark), and meetings (e.g., #hadoopsummit) that were unseen at t, continually entered into communities (e.g., #hadoop) at t + 1. This is the case for most of the communities (e.g., #iot, #cloud, #hr, #hadoop, #aws, #security) in the big data ecosystem.

The ecosystem also evolves through selective retention as communities grow, contract, and disappear from the big data ecosystem over time (Fig. 7). Growth appears to have been the de facto mechanism in the evolution of big data ecosystems, as big data as innovation has been widely diffused, new elements (e.g., #deeplearning, #yarn, #spark, #devops) have emerged, and existing elements have been copied. Many communities, including #iot, #hr, #aws and #agriculture, show this

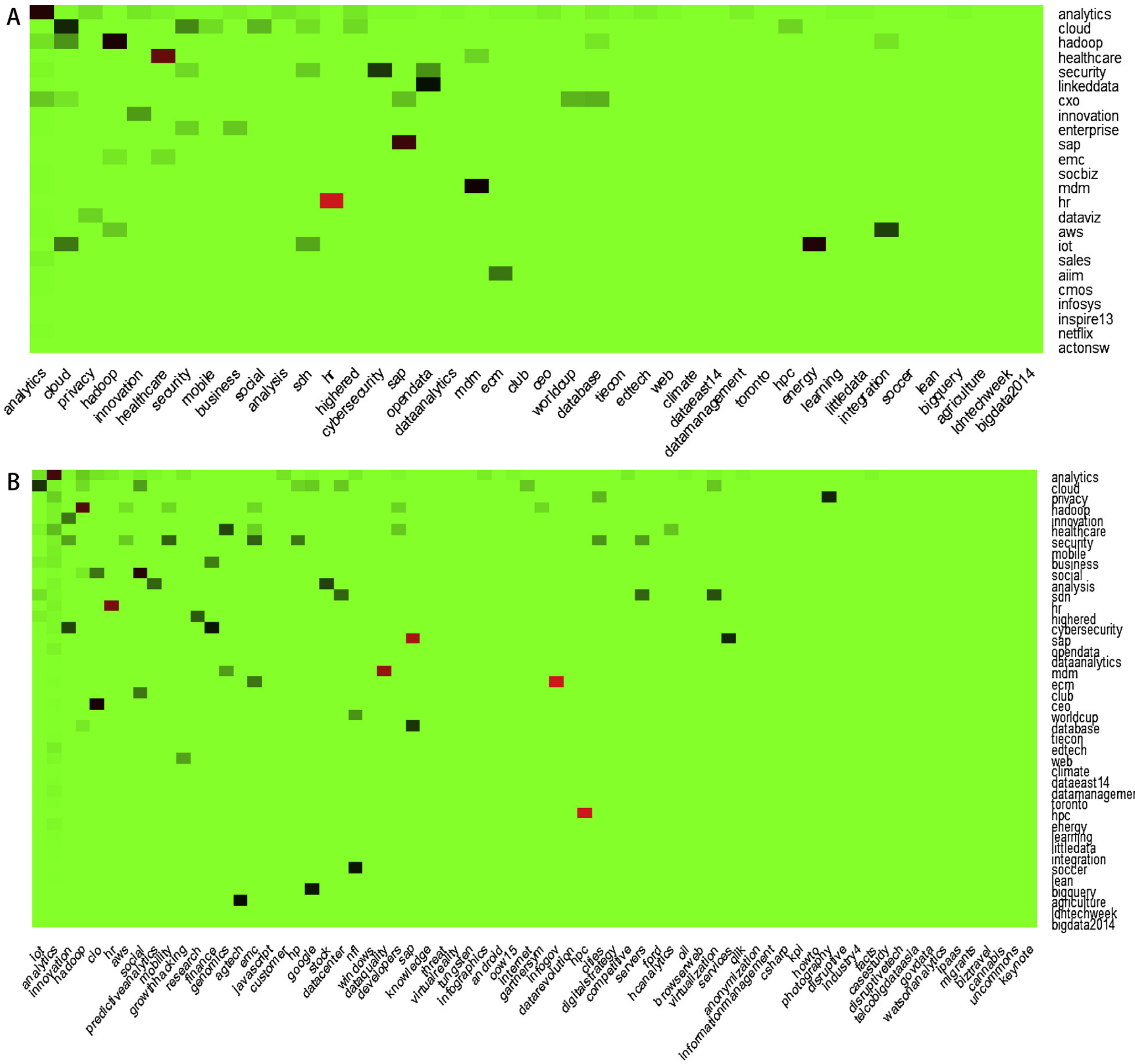

Fig. 5. (a) Community Similarity between 2013 (Y-axis) and 2014 (X-axis), (b) Community Similarity between 2014 (Y-axis) and 2015 (X-axis).

growth pattern. Nevertheless, the evolution is selective, as some communities become less prevalent over time. While this change may be rare in this fast-growing innovation ecosystem between 2013 and 2015, the importance of some communities has shrunk. For example, a community of #linkeddata, #semanticweb, #opendata, and #sparql (SPARQL Protocol and RDF Query Language) has become less visible over time. Some communities have faded away from the ecosystem. Almost all of these dead communities are small in size. Many event-based communities belong to this category. Professional meetings involve different groups of organizations, professionals, concepts, technologies, tools, and skills. These venues are necessary for innovation diffusion. They appear to be short-lived or to continue with new titles and organizers (Fig. 8).

6. Discussion

Recent studies of digital innovation have gone beyond individual or organizational technologies and have focused on such areas as digital infrastructure and product and service innovation (Barrett, Davidson, Prabhu, & Vargo, 2015; Um et al., 2015; Yoo et al., 2012). These studies have suggested resource heterogeneity, dynamic changes, and recombinant logic as key characteristics of digital innovation. That is, diverse elements, both social and technical, dynamically interact, and the digital innovation ecosystem emerges from such interaction and evolves over time. The boundary of the digital innovation ecosystem is fluid. In this line, past research has viewed big data as the latest case of digital innovation, which has evolved over time (Chae, 2015a; Chen, Chiang, & Storey, 2012; Yaqoob et al., 2016).

Using the proposed framework, this study has revealed a large number of elements that are diverse in form, including organizations, individuals, technologies, concepts, applications, regulations, professional meetings and associations, tools, and knowledge. This finding is similar to the results of some previous studies (e.g., Chae, 2015a; Yaqoob et al., 2016). In addition, the findings show that several consulting and data service companies and advanced users of big data technologies have been important members of this fast-growing ecosystem. Also, many new concepts (e.g., #analytics, #iot), technologies (e.g., #hadoop), applications (e.g., #healthcare), and infrastructures (e.g., #cloud) are strong drivers of the ecosystem. Security (e.g., #security), privacy (e.g., #privacy), and compliance (e.g., #compliance) have been consistently popular elements over the years. New roles (e.g., #chiefofdata, #datascientist) have emerged and specific knowledge

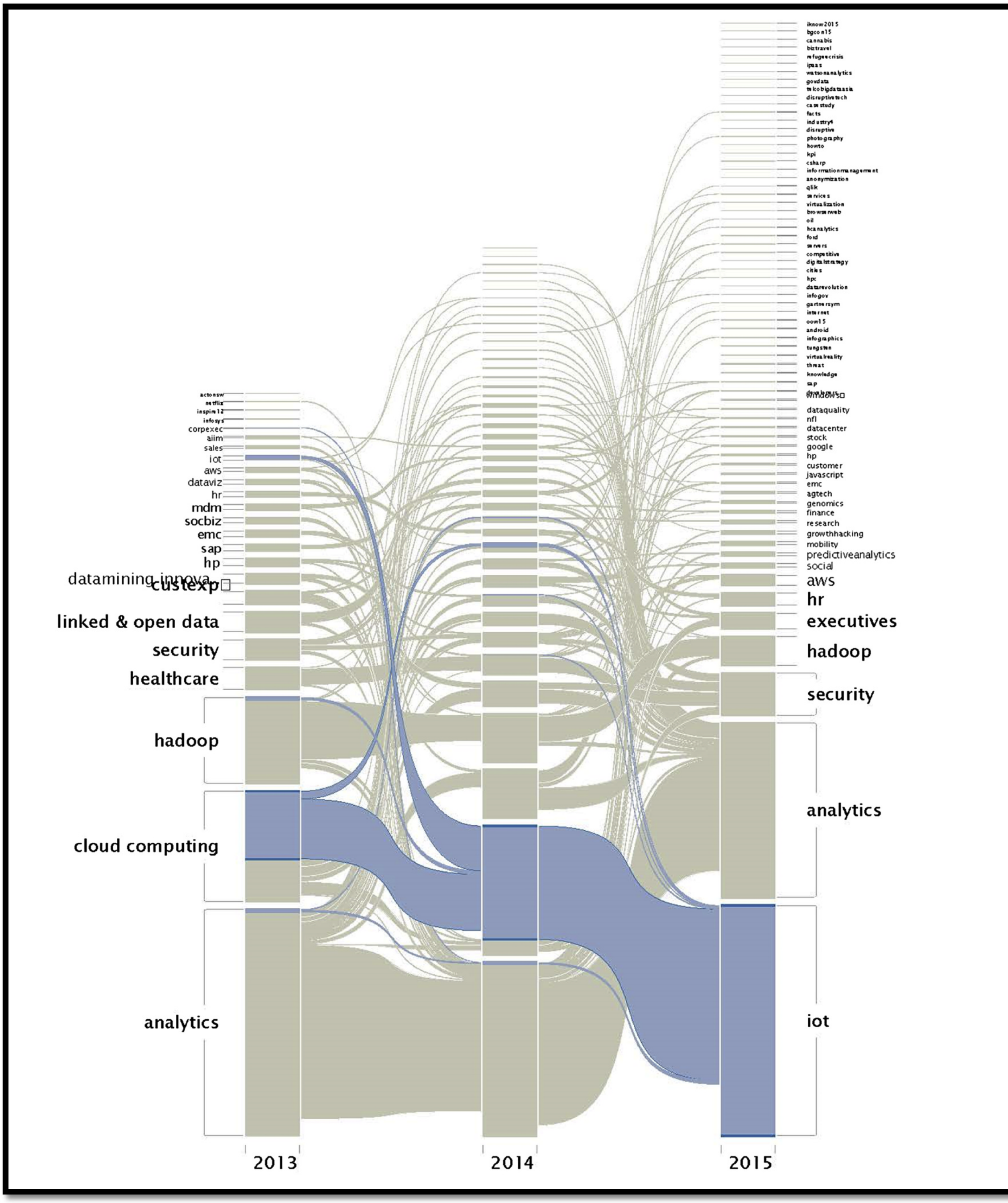

Fig. 6. Mapping evolutionary changes highlighting iot.

and skills (e.g., #statistics, #programming, #storytelling) are recognized in the ecosystem. Between 2013 and 2015, the elements in the ecosystem are tripled and the ecosystem becomes increasingly diverse.

The elements are also shown to be diverse in terms of their capacity, as indicated by the centrality scores (PageRank) (Bohlin et al., 2014) in the networks. In the early days (2013), traditional concepts (e.g., #businessintelligence, #bi) and technologies and tools (e.g., #excel, #sql) were still important parts of the ecosystem (Chen et al., 2012). In 2014 and 2015, other concepts, technologies, and tools have become more central in the ecosystem. In addition, in the early days of the ecosystem, consulting firms (e.g., #sap, #emc, #infosys) represent the majority of the elements in terms of organizations. In later years, other organizations, including service and consumer businesses, emerge in the ecosystem. This change is also apparent in other types of ecosystem elements (e.g., technologies, application areas).

In addition to their diversity in form and capacity, both those elements and their connections have evolved in the big data ecosystem. For example, Hadoop (#hadoop), which is a distributed big data analytics framework, was closely connected with the de facto data processing model called MapReduce (#mapreduce) by 2013 (e.g., Yaqoob et al., 2016), but this interaction has changed over time, as a new element (#spark) has emerged and replaced MapReduce. This type of information could be used to predict structural changes in the ecosystem and the popularity of technologies, applications, issues, and other elements.

The evolutionary changes in the digital ecosystem become clearer when the data are interpreted through a complexity lens (e.g., Yoo, 2015), and the mechanisms of variation and selective retention are

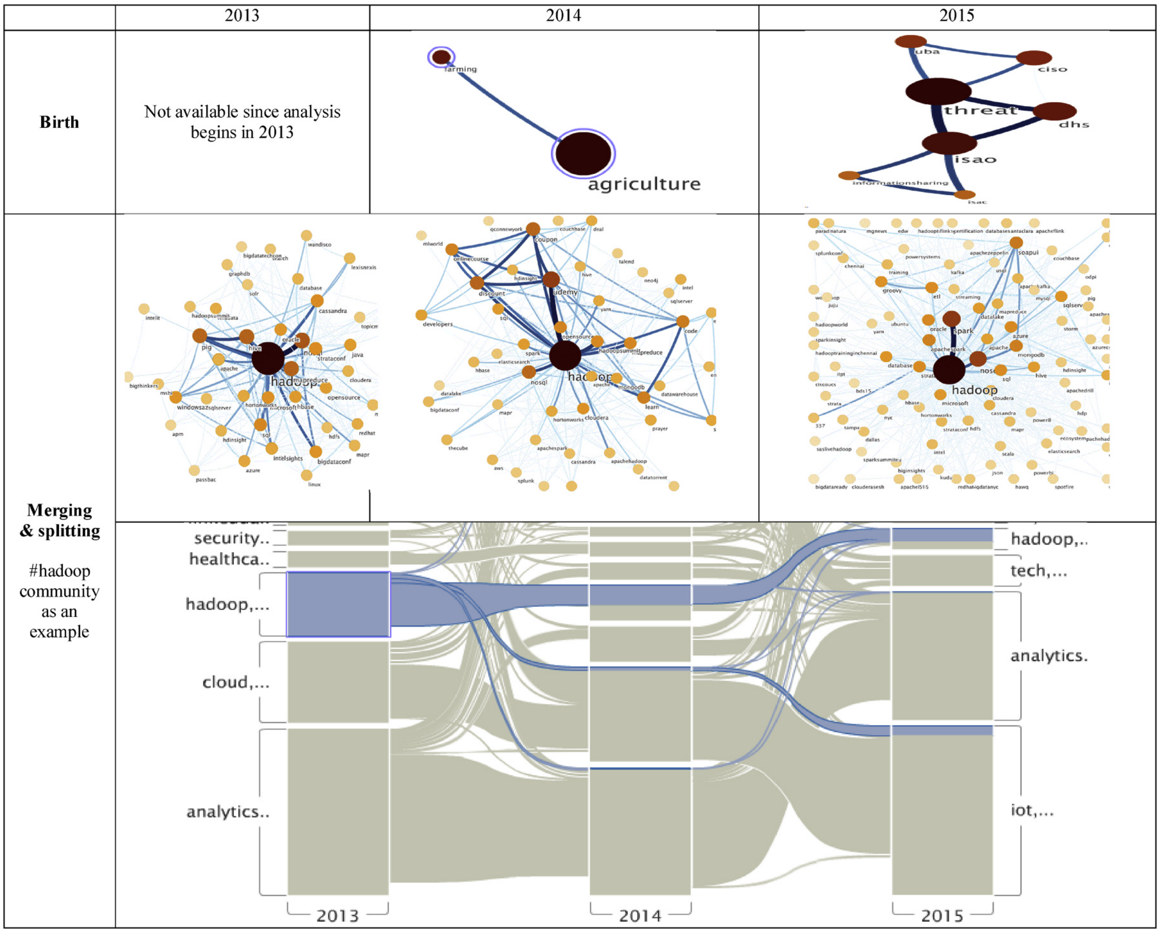

Fig. 7. Illustration of birth, merging, and splitting.

mapped through network science techniques like community analysis and alluvial diagrams. Communities have been merging and splitting, resulting in new communities and increasing the diversity of the ecosystem over time. New communities (e.g., #genomics, #climatechange, #agriculture, #datacenter) have also emerged from the interactions of those elements, which are previously unseen in the ecosystem. These elements come from other ecosystems (e.g., agriculture, biology) (e.g., Kauffman, 1995). The innovation has spread into multiple industries or fields. The ecosystem has been shown to be increasingly diverse through birth, merging, and splitting.

Also, communities grow, get smaller, or even fade away. Which pattern of selective retention is more popular than others in the big data ecosystem? There is more growth than contraction or death in the evolution of communities. A noticeable example is a network of concepts, organizations, technologies, and methodologies linked to #iot (Internet of Things). This community, comprised of diverse technologies, methodologies, and frameworks (e.g., #sensor, #microservices, #devops, #cisco, #fogcomputing), appears as one of the smallest in 2013, but takes the center position in 2015. This indicates that machine-generated data and relevant algorithmic techniques and tools are increasingly important in the ecosystem.

Among many elements, #analytics appears to have taken a unique position in the ecosystem. There are precedents, such as "business intelligence", of big data innovation. The concept #analytics makes big data innovation distinguishable from its precedents (Chen et al., 2012). While elements self-organize and there is no centralized design in big data innovation, #analytics appears to influence the emergence of new elements (e.g., #hranalytics, #webanalytics, #ibmanalytics) and communities (e.g., #predictiveanalytics, #watsonanalytics). This conceptual element and its community also remain very important, as many other elements attempt to be associated with #analytics. #analytics works more like a tagger (Holland, 1995) or a leader without authority (Schneider & Somers, 2006) in complex networks, which provides "direction" to elements and their interactions and also makes them distinguishable from others. In 2015, #iot seems to be taking over this role in the ecosystem. This emergence could potentially mean a major restructuring, in some aspects, in the big data ecosystem[2]. This change could have some significant implications, which would need further inquiry in the future.

6.1. Contribution to theory

Despite some recent studies going beyond individual and organizational technologies, macro level analysis of digital innovation is still rare in the digital innovation literature. Also, there are different theoretical frameworks for studying digital innovation in the literature, but most continue to be stage-based models and are not designed to explain the evolutionary and nonlinear pattern of the digital innovation ecosystem highlighted in this study. Recent commentaries have called for research of digital innovation using computational methods and evolutionary ontologies (Nambisan et al., 2017; Yoo, 2015). This study has responded to such a call by proposing a novel framework combining a complexity lens and a computational research methodology.

The complexity lens has offered a language to explain the continual emergence of communities and the big data ecosystem from the interaction among diverse elements. Two evolutionary mechanisms,

[2] The author thanks an anonymous reviewer for this insight.

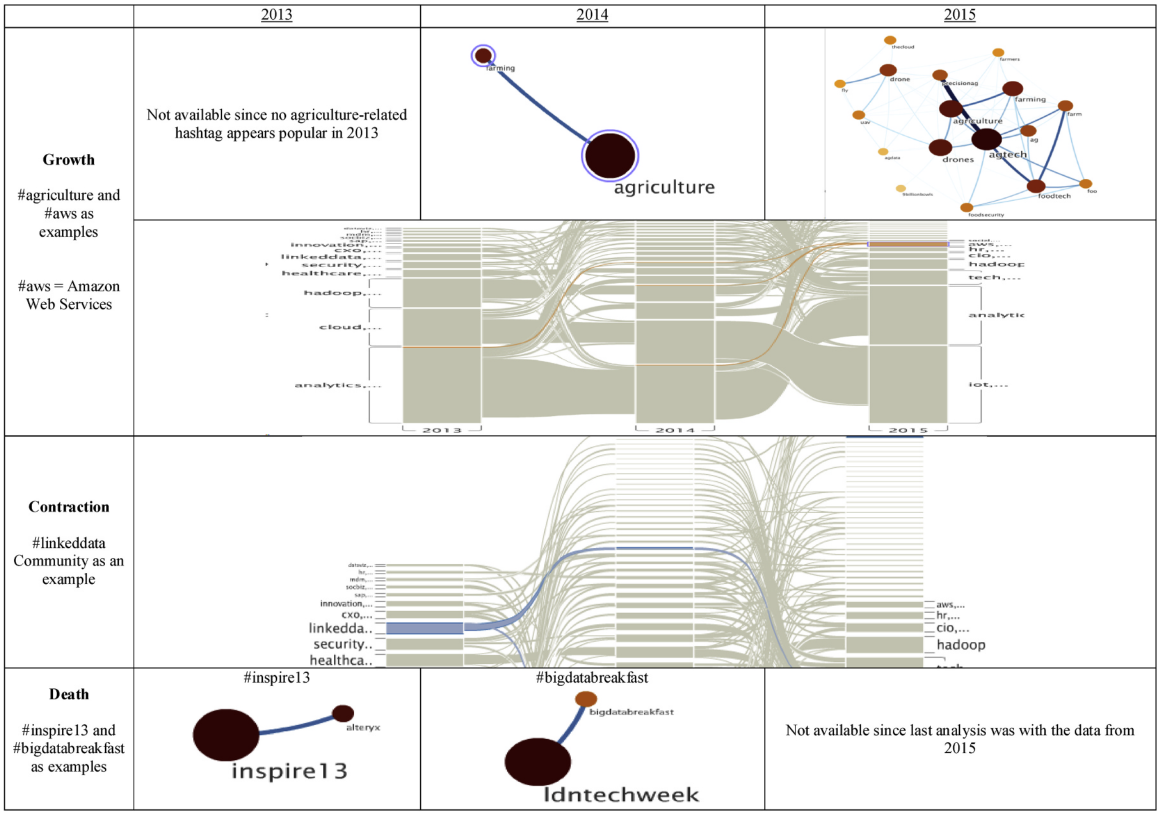

Fig. 8. Illustration of growth, contraction, and death.

variation and selective retention, are shown to be helpful in analyzing how the ecosystem has changed over time. Previous studies using a complexity lens tended to be conceptual or simulation-based. Our study has shown how network science methods can demonstrate the complexity view of digital innovation at the macro level. Specifically, community detection analysis with the map equation framework has offered a method to visualize the evolutionary and nonlinear patterns of how communities and the ecosystem emerge from the past and the interaction among diverse elements. Future studies would find the proposed framework useful for studying the evolution of different types of digital innovation.

6.2. Implications for practice

Digital innovation changes the way businesses and governments operate, and is also a driving force behind business and economic growth these days. It is increasingly important for managers and policy makers to track the emergence of digital innovation. As demonstrated in this study, however, digital innovation is a complex arrangement of diverse elements from different innovation ecosystems, and tracking its evolution is a daunting task for such practitioners. As a result, an understanding of the trends of a digital innovation largely relies on industry experts or organizations whose knowledge tends to be too expensive for individual analysts and policy makers to access.

In this vein, this study has provided a valuable tool in analyzing how a specific digital innovation is changing and what the important players in the ecosystem are. Specifically, the framework offers managers and policy makers a conceptual lens showing how a digital innovation is unfolding through variation and selective retention, where communities continually emerge from the interaction among existing and new elements. The network science techniques introduced in this study are practical enough for practitioners to apply to digital innovation in their own areas. The proposed conceptual lens and computational methods help them (1) identify diverse elements and their network positions in a respective digital innovation ecosystem and, (2) understand how different communities emerge from such elements and their interactions and, ultimately, how the ecosystem has evolved over time.

For example, managers and policy makers in the banking industry would investigate blockchain as digital innovation using digital trace data, evolutionary community detection analysis, and network visualization. This could reveal what technologies, organizations, applications, and issues have been central and peripheral actors in the blockchain ecosystem and, among them, which elements are closely related to banking and financial industry. This type of analysis can be done on a regular basis using computational tools, such as the map equation. Understanding the ecosystem would enable them to make timely and informed decisions in terms of blockchain-related resource allocations and new service developments.

7. Conclusion

This paper has proposed a framework relying on the notion of complex network as a theoretical lens and digital data and network science techniques for the empirical inquiry. The notion of complex networks offers a suitable conceptual lens to understand the emergence and evolution of digital innovation at the macro level. In the proposed approach, digital data is considered a key source of digital innovation research. Such a large dataset can be coded through text processing for computational network analysis. Computational techniques, such as community detection analysis and mapping network changes, are proven to be very helpful for understanding the emergence and evolution of the big data ecosystem. The findings have highlighted the evolution of digital innovation by two mechanisms, variation and selective retention, which are nonlinear and often unpredictable. The computational techniques introduced in the paper can be a useful tool to managers and policy makers as they understand the evolution of digital innovation and make timely and informed decisions at strategic levels, such as new service development and resource allocations. To this end, the proposed approach makes both theoretical and methodological contributions to digital innovation research and practice.

7.1. Limitations and future research directions

There are limitations in terms of data collection and analysis and opportunities for future research. First, the macro-level analysis of big data innovation could be enriched through more diversity in data sources and longer duration in data collection. This study uses Twitter data surrounding big data conversation with the assumption that this computational data collection would help capture some of the most popular elements (e.g., concepts, technologies, methodologies, skillsets, institutional contexts, economic actors) in big data as a digital innovation. However, rather than relying on a single source of digital data, additional data from other digital sources, such as electronic news articles, IT magazines, and web pages, could provide a more comprehensive picture of the big data ecosystem. Also, the duration of data collection could be longer, including Twitter data from as early as 2011. Google trends indicate that "big data" appears as a search term as early as 2011.

Second, this study uses Twitter hashtags to construct complex networks of heterogeneous elements and their connections. It is possible there is a large amount of big data conversation in Twitter without using hashtags, for example, "deep learning" and "neural network" rather than #deeplearning and #neuralnetwork. To capture "deep learning" and "neural network" as elements of the big data ecosystem, future research can consider advanced natural language processing techniques and tools, such as named entity recognition (Li, Sun, Weng, & He, 2015), which identifies names of entities, such as organizations, products, and people. These techniques can be adapted to extract important entities, not necessarily in the hashtag format, from digital data. This information can then be converted to automatically represent different types of elements (e.g., concept, organization, product, technology, job title) in complex networks, and used to construct multimodal networks. This can potentially enrich the macro-level analysis of digital innovation.

Acknowledgement

The author acknowledges the summer research support from the College of Business at Kansas State University, Manhattan KS, USA.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems, 17* Article 3.
Aswani, R., Kar, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2018). Search engine marketing is not all gold: Insights from Twitter and SEOClerks. *International Journal of Information Management, 38*, 107–116.
Axelrod, R., & Cohen, M. (2000). *Harnessing complexity.* Basic Books.
Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. (2015). Service innovation in the digital age: Key contributions and future directions. *MIS Quarterly, 39*, 135–154.
Bech, M. L., Bergstrom, C. T., Rosvall, M., & Garratt, R. J. (2015). Mapping change in the overnight money market. *Physica A Statistical Mechanics and Its Applications, 424*, 44–51.
Bohlin, L., Edler, D., Lancichinetti, A., & Rosvall, M. (2014). *Community detection and visualization of networks with the map equation framework. Measuring scholarly impact.* Springer3–34.
Campbell, D. (1960). Blind varieties and selective retention in creative thought and other processes. *Psychological Review, 67*, 380–400.
Chae, B. K. (2014). A complexity theory approach to IT-Enabled services (IESs) and service innovation: Business analytics as an illustration of IES. *Decision Support Systems, 57*, 1–10.
Chae, B. K. (2015a). Big data and IT-Enabled services: Ecosystem and coevolution. *IT Professional, 17*, 20–25.
Chae, B. K. (2015b). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics, 165*, 247–259.
Chang, Y.-C., Ku, C.-H., & Chen, C.-H. (2017). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management*.
Chen, H., Chiang, D., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*, 1–24.
Chiles, T. H., Meyer, A., & Hench, T. J. (2004). Organizational emergence: The origin and transformation of branson missouri's musical theaters. *Organization Science, 15*, 499–519.
de Camargo Fiorini, P., Roman Pais Seles, B. M., Chiappetta Jabbour, C. J., Barberio Mariano, E., & de Sousa Jabbour, A. B. L. (2018). Management theory and big data literature: From a review to a research agenda. *International Journal of Information Management, 43*, 112–129.
Fichman, R. G., Dos Santos, B. L., & Zheng, Z. E. (2014). Digital innovation as a fundamental and powerful concept in the information systems curriculum. *MIS Quarterly, 38*, 329–353.
Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*, 75–174.
Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports, 659*, 1–44.
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144.
Garud, R., Tuertscher, P., & Van de Ven, A. H. (2013). Perspectives on innovation processes. *The Academy of Management Annals, 7*, 775–819.
Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*, 7821–7826.
Goes, P. B. (2014). Big data and IS research. *MIS Quarterly, 38* iii-viii.
Grover, P., Kar, A. K., & Davies, G. (2018). "Technology enabled Health"–Insights from twitter analytics with a socio-technical perspective. *International Journal of Information Management, 43*, 85–97.
Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2018). Polarization and acculturation in US Election 2016 outcomes – Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change*.
Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management, 42*, 78–89.
Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98–115.
Holland, J. (1995). *Hidden order.* Cambridge: Perseus Books.
Holland, J. H. (2014). *Complexity: A very short introduction.* Oxford University Press.
Holme, P. (2015). Modern temporal network theory: A colloquium. *The European Physical Journal B, 88*, 234.
Hu, H., Huang, T., Zeng, Q., & Zhang, S. (2016). The role of institutional entrepreneurship in building digital ecosystem: A case study of Red Collar Group (RCG). *International Journal of Information Management, 36*, 496–499.
Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers, 20*, 531–558.
Kauffman, S. (1995). *At home in the universe: The search for laws of self-organization and complexity.* New York: Oxford University Press.
Li, C., Sun, A., Weng, J., & He, Q. (2015). Tweet segmentation and its application to named entity recognition. *IEEE Transactions on Knowledge and Data Engineering, 27*, 558–570.
Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital world. *Management Information Systems Quarterly, 41*, 223–238.
Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*.
Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*, 664–667.
Pan, Y., Li, D.-H., Liu, J.-G., & Liang, J.-Z. (2010). Detecting community structure in complex networks via node similarity. *Physica A Statistical Mechanics and Its Applications, 389*, 2849–2857.
Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communities in networks. *Notices of the AMS, 56*, 1082–1097.
Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *International Journal of Information Management, 38*, 187–195.
Rathore, A. K., Kar, A. K., & Ilavarasan, P. V. (2017). Social media analytics: Literature review and directions for future research. *Decision Analysis, 14*, 229–249.
Rossetti, G., & Cazabet, R. (2018). Community discovery in dynamic networks: A survey. *ACM Computing Surveys (CSUR), 51*, 35.
Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics, 178*, 13–23.
Rosvall, M., & Bergstrom, C. T. (2010). Mapping change in large networks. *PloS One, 5*, e8694.
Sagiroglu, S., & Sinanc, D. (2013). *Big data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on (Pp. 42-47): IEEE*.
Schneider, M., & Somers, M. (2006). Organizations as complex adaptive systems: Implications of Complexity Theory for leadership research. *The Leadership Quarterly, 17*, 351–365.
Scott, W. (2008). *Institutions and organizations* (3rd ed.). London: Sage Publications.
Shi, Z., Rui, H., & Whinston, A. (2014). Content sharing in a social broadcasting environment: Evidence from twitter. *MIS Quarterly, 38*, 123–142.
Steyvers, M., & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*, 41–78.
Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management, 39*, 156–168.

Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*, 268–276.
Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science, 60*, 1452–1469.
Tiwana, A. (2015). Evolutionary competition in platform ecosystems. *Information Systems Research*.
Tria, F., Loreto, V., Servedio, V. D. P., & Strogatz, S. H. (2014). The dynamics of correlated novelties. *Scientific Reports, 4*.
Um, S., Yoo, Y., & Wattal, S. (2015). The evolution of digital ecosystems: A case of WordPress from 2004 to 2014. *International Conference on Information Systems*.
Wang, P. (2009). Popular concepts beyond organizations: Exploring new dimensions of information technology innovations. *Journal of the Association for Information Systems, 10*, 1–30.
Wang, X., Kuzmickaja, I., Stol, K.-J., Abrahamsson, P., & Fitzgerald, B. (2014). Microblogging in open source software development: The case of Drupal and Twitter. *IEEE Software,* 72–80.
Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., et al. (2016). Big data: From beginning to future. *International Journal of Information Management, 36*, 1231–1247.
Yoo, Y. (2012). Digital materiality and the emergence of an evolutionary science of the artificial. In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.). *Materiality and organizing: Social interaction in a technological world* (pp. 134–154). Oxford University Press.
Yoo, Y. (2015). It is not about size: A further thought on big data. *Journal of Information Technology, 30*, 63–65.
Yoo, Y., Boland, R., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science, 23*, 1398–1408.
Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research, 21*, 724–735.